

Symmetry of Information: A Closer Look

Marius Zimand ^{*}.

Department of Computer and Information Sciences, Towson University, Baltimore,
MD, USA

Abstract. Symmetry of information establishes a relation between the information that x has about y (denoted $I(x : y)$) and the information that y has about x (denoted $I(y : x)$). In classical information theory, the two are exactly equal, but in algorithmical information theory, there is a small excess quantity of information that differentiates the two terms, caused by the necessity of packaging information in a way that makes it accessible to algorithms. It was shown in [Zim11] that in the case of strings with simple complexity (that is the Kolmogorov complexity of their Kolmogorov complexity is small), the relevant information can be packed in a very economical way, which leads to a tighter relation between $I(x : y)$ and $I(y : x)$ than the one provided in the classical symmetry-of-information theorem of Kolmogorov and Levin. We give here a simpler proof of this result. This result implies a van Lambalgen-type theorem for finite strings and plain complexity: If x is c -random and y is c -random relative to x , then xy is $O(c)$ -random. We show that a similar result holds for prefix-free complexity and weak-K-randomness.

1 Introduction

In classical information theory, the information that X has about Y is equal to the information that Y has about X , i.e.,

$$I(X : Y) = I(Y : X).^1$$

In algorithmical information theory, there exists a similar relation, but it is less perfect:

$$I(x : y) \approx I(y : x).$$

In this paper we take a closer look at the level of approximation indicated by “ \approx .” In the line above, x and y are binary strings of lengths n_x and, respectively, n_y , and $I(x : y)$, the information in x about y , is defined by $I(x : y) = C(y \mid n_y) - C(y \mid x, n_y)$, where $C(\cdot)$ is the plain Kolmogorov complexity. The precise form of

^{*} The author is supported in part by NSF grant CCF 1016158. Most of the results have been presented at International Workshop on Theoretical Computer Science, Feb 21-24, 2012, Auckland, New Zealand. URL: <http://triton.towson.edu/~mzimand>

¹ X and Y are random variables and $I(X : Y) = H(Y) - H(Y \mid X)$, where H is Shannon entropy.

the equation is given by the classical Kolmogorov-Levin theorem [ZL70], which states that

$$|I(x : y) - I(y : x)| = O(\log n_x + \log n_y). \quad (1)$$

This is certainly an important result, but in some situations, it is not saying much. Suppose, for example, that x has very little information about y , say, $I(x : y)$ is a constant (not depending on x and y). What does this tell about $I(y : x)$? Just from the definition and ignoring $I(x : y)$, $I(y : x)$ can be as large as $n_x - O(1)$, and, somewhat surprisingly, given that $I(x : y) = O(1)$, $I(y : x)$ can still have the same order of magnitude. As an example, consider y a c -random string (meaning $C(y | n_y) \geq n_y - c$), whose length n_y is also c -random. Let x be the binary encoding of n_y . Then $I(x : y) = C(y | n_y) - C(y | x, n_y) = O(1)$, but $I(y : x) = C(x | n_x) - C(x | y, n_x) = n_x - O(1)$. This example shows that even though, for some strings, the relation (1) is trivial, it is also tight.

In many applications, the excess term $O(\log n_x + \log n_y)$ does not hurt too much. But sometimes it does. For example, it is the reason why the Kolmogorov-Levin Theorem does not answer the following basic “direct product” question: If x and y are c -random n -bit strings and $I(x : y) \leq c$, does it follow that xy is $O(c)$ -random?

The answer is positive (providing the finite analog of the van Lambalgen theorem). It follows from a recent result of the author [Zim11], which establishes a more precise symmetry-of-information relation for strings with *simple complexity*, i.e., for strings x such that $C^{(2)}(x | n_x)$ is small, say, bounded by a constant.² Note that all random strings have simple complexity, and that there are other types of strings with simple complexity as well. The main result in [Zim11] implies that if x and y are strings of equal length that have simple complexity bounded by c and $I(x : y) \leq c$, then $I(y : x) \leq O(c)$.

The proof method in [Zim11] is based on randomness extraction. Alexander Shen (personal communications) has observed that in fact a proof of the above result in [Zim11] can be obtained via the standard method used in the proof of Kolmogorov-Levin Theorem (see [DH10, LV08]). We present here such a proof.

We slightly extend the symmetry-of-information relation from [Zim11] to strings x and y that may have different lengths. We prove the following (with the convention that $\log 0 = 0$):

Theorem 1 (Main Theorem). *For all strings x and y ,*

$$I(y : x) \leq I(x : y) + O(\log I(x : y)) + O(\log |n_x - n_y|) + \delta(x, y),$$

where $\delta(x, y) = O(C^{(2)}(x | n_x) + C^{(2)}(y | n_y))$.

Thus, for strings x and y with simple complexity, $I(y : x) \leq I(x : y) + O(\log I(x : y)) + O(\log |n_x - n_y|)$.

As mentioned, Theorem 1 implies an analog of the van Lambalgen Theorem for finite strings and plain Kolmogorov complexity. Recall that van Lambalgen

² The notation $C^{(2)}(x | n_x)$ is a shorthand for $C(C(x | n_x) | n_x)$.

Theorem [vL87] states that for infinite sequences x and y the following two statements are equivalent:

- (1) x is random and y is random relative by x ,
- (2) $x \oplus y$ is random.

(Here, random means Martin-Löf random, and $x \oplus y$ is the infinite sequence obtained by alternating the bits of x and y). Theorem 1 implies immediately the following.

Theorem 2 ([Zim11]). *Let x and y be two strings of length n such that x is c -random and y is c -random relative to x . Then xy is $O(c)$ -random.*

Formally, if $C(x | n) \geq n - c$, $C(y | x) \geq n - c$, then $C(xy | 2n) \geq 2n - O(c)$.

(The constant in the $O(\cdot)$ notation depends only on the universal Turing machine.)

Is there a van Lambalgen Theorem analog for prefix-free complexity? To adress this question, we note that there are two notions of randomness for the prefix-free complexity K , weak K -randomness, and strong K -randomness (see [DH10]). An n -bit string x is

- (1) weakly c -(K -random) if $K(x | n) \geq n - c$, and
- (2) strongly c -(K -random) if $K(x | n) \geq n + K(n) - c$.

For weak K -randomness, we prove in Section 4 the following analog of the van Lambalgen Theorem. For strong K -randomness, the question remains open.

Theorem 3. *Let x and y be two strings of length n such that x is weakly c -(K -random) and y is weakly c -(K -random) relative to x . Then xy is weakly $O(c)$ -(K -random).*

Formally, if $K(x | n) \geq n - c$, $K(y | x) \geq n - c$, then $K(xy | 2n) \geq 2n - O(c)$.

(The constant in the $O(\cdot)$ notation depends only on the universal Turing machine.)

1.1 Proof techniques

The proofs of Symmetry of Information Theorems have a combinatorial nature. To fix some ideas, let us first sketch the proof of the classical Kolmogorov-Levin Theorem which establishes relation (1). The setting of the theorem, as well as for the rest of our discussion, is as follows: x and y are binary strings of lengths n_x and, respectively, n_y . An easy observation (see Section 2.2), shows that a relation between $I(y : x)$ and $I(x : y)$ can be deduced from a “chain rule:” $C(xy | n_x, n_y) \geq C(x | n_x) + C(y | x, n_y) - (\text{small term})$. In our case, to obtain (1), it is enough to show that

$$C(xy | n_x, n_y) \geq C(x | n_x) + C(y | x, n_y) - O(\log n_x + \log n_y). \quad (2)$$

Let us consider a $2^{n_x} \times 2^{n_y}$ table with rows indexed by $u \in \{0, 1\}^{n_x}$ and columns indexed by $v \in \{0, 1\}^{n_y}$. Let $t = C(xy | n_x, n_y)$. We assign boolean values to the cells of the table as follows. The (u, v) -cell in the table is 1 if $C(uv | n_x, n_y) \leq t$

and it is 0 otherwise. The number of cells equal to 1 is less than 2^{t+1} , because there are only 2^{t+1} programs of length $\leq t$. Let m be such that the number of 1's in the x row is in the interval $(2^{m-1}, 2^m]$. Note that, given x, n_y and t , we can enumerate the 1-cells in the x -row and one of them is the (x, y) cell. It follows that

$$C(y \mid x, n_y) \leq m + O(\log t). \quad (3)$$

Now consider the set of rows that have at least 2^{m-1} 1s. The number of such rows is at most $2^{t+1}/2^{m-1} = 2^{t-m+2}$. We can effectively enumerate these rows if we are given n_x, n_y, m and t . Since the x row is one of these rows, it follows that

$$C(x \mid n_x) \leq t - m + O(\log n_y + \log m + \log t). \quad (4)$$

Adding equations (3) and (4) and keeping into account that $\log m \leq \log n_y$ and $\log t \leq O(\log n_x + \log n_y)$, the relation (2) follows.

A careful inspection of the proof reveals that the excess term $O(\log n_x + \log n_y)$ is caused by the fact that the enumerations involved in describing y and x need to know m (which is bounded by $C(y \mid n_y)$) and $t = C(xy \mid n_x, n_y)$. In case x and y are c -random, it is more economical to use randomness deficiencies. In particular instead of using $t = C(xy \mid n_x, n_y)$, we can do a similar argument based on the randomness deficiency of xy , i.e., $w = (n_x + n_y) - C(xy \mid n_x, n_y)$. This is an observation of Chang, Lyuu, Ti and Shen [CLTS09], who attribute it to folklore. For strings with simple complexity, it is advantageous to express t as $t = C(x \mid n_x) + C(y \mid x) - w$ because the first two terms have a short description. This yields a proof of the *Main Theorem* which we present in Section 3.

2 Preliminaries

2.1 Notation and background on Kolmogorov complexity

The Kolmogorov complexity of a string x is the length of the shortest effective description of x . There are several versions of this notion. We use here mainly the *plain complexity*, denoted $C(x)$, and the *conditional plain complexity* of a string x given a string y , denoted $C(x \mid y)$, which is the length of the shortest effective description of x given y . The formal definitions are as follows. We work over the binary alphabet $\{0, 1\}$. A string is an element of $\{0, 1\}^*$. If x is a string, n_x denotes its length. Let M be a Turing machine that takes two input strings and outputs one string. For any strings x and y , define the *Kolmogorov complexity* of x conditioned by y with respect to M , as $C_M(x \mid y) = \min\{|p| \mid M(p, y) = x\}$. There is a universal Turing machine U with the following property: For every machine M there is a constant c_M such that for all x , $C_U(x \mid y) \leq C_M(x \mid y) + c_M$. We fix such a universal machine U and dropping the subscript, we write $C(x \mid y)$ instead of $C_U(x \mid y)$. We also write $C(x)$ instead of $C(x \mid \lambda)$ (where λ is the empty string). If n is a natural number, $C(n)$ denotes the Kolmogorov complexity of the binary representation of n . We use $C^{(2)}(x \mid n_x)$ as a shorthand for $C(C(x \mid n_x) \mid n_x)$. For two strings x and y , the information in x about y is denoted $I(x : y)$ and is defined as $I(x : y) = C(y \mid n_y) - C(y \mid x, n_y)$.

Prefix-free complexity K is defined in a similar way, the difference being that the universal machine is required to be prefix-free.

In this paper, the constant hidden in the $O(\cdot)$ notation only depends on the universal Turing machine.

In this paper, by convention $\log 0 = 0$.

2.2 Symmetry of Information and the Chain Rule

All the forms of the Symmetry of Information that we are aware of have been derived from the *chain rule* and this paper is no exception. The chain rule states that $C(xy \mid n_x, n_y) \approx C(x \mid n_x) + C(y \mid x, n_y)$. For us it is of interest to have an accurate estimation of the “ \approx ” relation. It is immediate to see that $C(xy \mid n_x, n_y) \leq C(y \mid n_y) + C(x \mid y, n_x) + 2C^{(2)}(y \mid n_y) + O(1)$. If we show a form of the converse inequality

$$C(xy \mid n_x, n_y) \geq C(x \mid n_x) + C(y \mid x, n_y) - (\text{small term}), \quad (5)$$

then we deduce that $C(x \mid n_x) - C(x \mid y, n_x) \leq C(y \mid n_y) - C(y \mid x, n_y) + 2C^{(2)}(y \mid n_y) + (\text{small term})$, i.e.,

$$I(y : x) \leq I(x : y) + 2C^{(2)}(y \mid n_y) + (\text{small term}).$$

Thus our efforts will be directed to establishing forms of the relation (5) in which (small term) is indeed small.

3 Proof of the Main Theorem

We demonstrate Theorem 1, using the standard proof method of the Kolmogorov-Levin Theorem.

Let $t_x = C(x \mid n_x)$, $t_y = C(y \mid x, n_y)$ and $t = C(xy \mid n_x, n_y)$. We take $w = t_x + t_y - t$, which is the term called (*small term*) in equation (5), and plays the role of randomness deficiency. Our goal is to show that

$$w = O(\log I(x : y)) + O(\log |n_x - n_y|) + \delta(x, y),$$

from which the Main Theorem follows (using the discussion in Section 2.2).

We assume that $w > 0$, otherwise there is nothing to prove. We will need to handle information $(|n_x - n_y|, t_x, t_y, w, b)$, where b is a bit that indicates if $n_x > n_y$ or not. Let Λ be a string encoding in a self-delimiting way this information, and let λ be the length of Λ . Note that

$$\lambda \leq 2 \log w + O(C^{(2)}(x \mid n_x) + C^{(2)}(y \mid n_y) + \log I(x : y) + \log |n_x - n_y|).$$

We build a $2^{n_x} \times 2^{n_y}$ boolean table, with rows and columns indexed by the strings in $\{0, 1\}^{n_x}$ and, respectively, $\{0, 1\}^{n_y}$, and we set the value of cell (u, v)

to be 1 if $C(uv \mid n_x, n_y) \leq t$, and to be 0 otherwise. Let S be the set of 1-cells, and S_u be the set of 1-cells in row u . Note that

$$|S| \leq 2^{t+1}.$$

Let m be defined as $2^{m-1} < |S_x| \leq 2^m$. We take F to be the set of “fat” rows, i.e., the set of rows having more than 2^{m-1} 1-cells. We have $|F| < |S|/2^{m-1} \leq 2^{t-m+2}$. Note that x is in F , and that the elements of F can be effectively enumerated given the information A and m . It follows that, given A and m , the string x can be described by its index in the enumeration of F . This index can be written on *exactly* $t - m + 2$ bits, so that knowing t which can be deduced from A , we can reconstruct m . It follows that

$$C(x \mid n_x, A) \leq t - m + 2 + O(1). \quad (6)$$

Next we note that y is in S_x and the elements of S_x can be enumerated given x and A . It follows that y can be described by its index in the enumeration of S_x . We obtain

$$C(y \mid x, A) \leq m + O(1). \quad (7)$$

From Equations (6) and (7), we conclude that

$$C(x \mid n_x, A) + C(y \mid x, A) \leq t + O(1) = t_x + t_y - w + O(1).$$

Note that $t_x = C(x \mid n_x) \leq C(x \mid n_x, A) + O(\lambda)$ and $t_y = C(y \mid x, n_y) \leq C(y \mid x, A) + O(\lambda)$. We obtain $t_x + t_y \leq t_x + t_y - w + O(\lambda)$. Taking into account the estimation for λ , we have $w - O(\log w) = O(C^{(2)}(x \mid n_x) + C^{(2)}(y \mid n_y) + \log I(x : y) + \log |n_x - n_y|)$, from which, $w = O(C^{(2)}(x \mid n_x) + C^{(2)}(y \mid n_y) + \log I(x : y) + \log |n_x - n_y|)$, as desired.

4 A van Lambalgen Theorem for prefix-free complexity

In this section we prove Theorem 3.

We introduce first the following notation: \bar{d} is a self delimited encoding of the integer d . We can have $|\bar{d}| \leq 2 \log d + 4$.

For a string u , let $A_u(d) = \{v \in \{0, 1\}^n \mid K(uv \mid n) \leq 2n - d\}$, and let $F(d) = \{u \in \{0, 1\}^n \mid |A_u(d)| \geq 2^{n-d}\}$.

Consider the following program p_1 , which has x of length n on the oracle tape (as conditional information):

On input the pair of positive integers (d, i) , encoded as $\bar{d} \text{ bin}(i)$, check if i is written on $n - d$ bits. If not, diverge. If yes, enumerate strings u of length n such that $K(xu \mid n) \leq 2n - 2d$. Output the i -th string in the enumeration.

Note that the domain of p_1 is prefix-free. Let c_1 be the length of program p_1 .

Let p_2 be the following program, which has n on the oracle tape (as conditional information):

On input the pair of positive integers (d, i) , encoded as $\overline{d} \text{ bin}(i)$, check if i is written on $n - d$ bits. If not, diverge. If yes, enumerate the elements of $F(d)$ and output the i -th one.

Program p_2 is prefix-free. Let c_2 be its length.

Let $d_0 = \max\{2(c + 4 + c_1), 2(c + 4 + c_2)\}$.

If there is some i such that $p_1(d_0, i)$ outputs u , then $K(u \mid x) \leq n - d_0 + 2 \log d_0 + 4 + c_1 < n - c$. So, there is no i such that $p_1(d_0, i)$ outputs y . This can happen only if either

(a) $K(xy \mid n) > 2n - 2d_0$, or

(b) $K(xy \mid n) \leq 2n - 2d_0$ but there are 2^{n-d_0} strings u which are enumerated before y .

If (a), we are done.

We show that (b) is impossible.

Indeed, if (b) holds, then $|A_x(d_0)| \geq 2^{n-d_0}$ and thus x is in $F(d_0)$. Also $|F(d_0)| \leq 2^{2n-2d_0}/2^{n-d_0} = 2^{n-d_0}$. It follows that for some i , program p_2 on input (d_0, i) outputs x . But then $K(x \mid n) < n - d_0 + 2 \log d_0 + 4 + c_2 < n - c$, which is a contradiction.

5 Final remarks

For symmetry of information, the argument from [Zim11] based on randomness extractors, a concept introduced and studied intensively in computational complexity, can be replaced by the simpler and more direct argument suggested by Sasha Shen and presented here. Nevertheless, we still think that extractors play a role in Kolmogorov complexity theory. They have been used in Kolmogorov complexity extraction (see [FHP⁺11, HPV11, Mus12, Zim10d, Zim09], and the survey paper [Zim10a]) and, of course, this is not surprising given the analogy between Kolmogorov complexity extraction and randomness extraction. But they have also been used for obtaining results that are not immediately connected to randomness extraction, such as counting results regarding dependent and independent strings [Zim10b], and independence amplification [Zim10c].

6 Acknowledgements

I am grateful to Alexander Shen who has noticed that the main theorem can be obtained with the method used in the standard proof of the Kolmogorov-Levin Theorem.

References

- CLTS09. C. Chang, Y. Lyuu, Y. Ti, and A. Shen. Sets of k -independent sets. *International Journal of Foundations of Computer Science*, 2009. To appear.
- DH10. R. Downey and D. Hirschfeldt. *Algorithmic randomness and complexity*. Springer Verlag, 2010.

- FHP⁺11. L. Fortnow, John M. Hitchcock, A. Pavan, N. V. Vinodchandran, and F. Wang. Extracting Kolmogorov complexity with applications to dimension zero-one laws. *Inf. Comput.*, 209(4):627–636, 2011.
- HPV11. J. M. Hitchcock, A. Pavan, and N. V. Vinodchandran. Kolmogorov complexity in randomness extraction. *TOCT*, 3(1):1, 2011.
- LV08. M. Li and P. Vitanyi. *An introduction to Kolmogorov complexity and its applications*. Springer-Verlag, 2008. 3rd edition. 1st edition in 1993.
- Mus12. Daniil Musatov. Space-bounded Kolmogorov extractors. *CoRR*, abs/1203.3674, 2012.
- vL87. M. van Lambalgen. Random sequences, Ph.D. dissertation. Technical report, University of Amsterdam, 1987.
- Zim09. M. Zimand. Extracting the Kolmogorov complexity of strings and sequences from sources with limited independence. In *Proceedings 26th STACS, Freiburg, Germany*, February 26–29 2009.
- Zim10a. M. Zimand. Possibilities and impossibilities in Kolmogorov complexity extraction. *SIGACT News*, 41(4):74–94, December 2010.
- Zim10b. Marius Zimand. Counting dependent and independent strings. In *MFCS*, volume 6281 of *Lecture Notes in Computer Science*, pages 689–700. Springer, 2010.
- Zim10c. Marius Zimand. Impossibility of independence amplification in Kolmogorov complexity theory. In *MFCS*, volume 6281 of *Lecture Notes in Computer Science*, pages 701–712. Springer, 2010.
- Zim10d. Marius Zimand. Two sources are better than one for increasing the Kolmogorov complexity of infinite sequences. *Theory Comput. Syst.*, 46(4):707–722, 2010.
- Zim11. M. Zimand. Symmetry of information and bounds on nonuniform randomness extraction via Kolmogorov complexity. In *Proceedings 26th IEEE Conference on Computational Complexity*, June 2011. To appear.
- ZL70. A. Zvonkin and L. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematical Surveys*, 25(6):83–124, 1970.